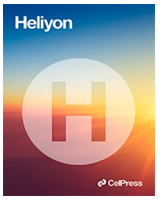
Heliyon 10 (2024) e40491

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Identification of biomarkers related to *Escherichia coli* infection for the diagnosis of gastrointestinal tumors applying machine learning methods

Tingting Ge, Wei Wang, Dandan Zhang, Xubo Le, Lumei Shi *

Department of Clinical Laboratory, Beilun People's Hospital, Ningbo, 315800, China

ARTICLE INFO

Keywords:
Gastrointestinal tumors
Escherichia coli
Biomarker
Machine learning
Diagnosis

ABSTRACT

Background: *Escherichia coli* (*E. coli*) is a part of normal gastrointestinal microbiota but it could also cause human gastrointestinal diseases. Understanding the mechanism of *E. coli* in the progression of gastrointestinal tumors can provide novel prevention and treatment strategies for gastrointestinal tumors.

Methods: The *E. coli* infection score was calculated by single sample GSEA (ssGSEA). Weighted correlation network analysis (WGCNA) and differentially expressed genes (DEGs) analysis were used to identify genes related to *E. coli* infection in gastrointestinal tumors. Hub genes were selected by machine learning methods to establish a diagnostic model. The diagnostic performance of the model was evaluated by the area under the receiver operating characteristic (ROC) curve (AUC) and validated in three external datasets. After determining the biomarkers, immune infiltration analysis and GSEA were further performed. The mRNA expressions of the biomarkers in stomach adenocarcinoma (STAD) cells and the invasion and migration of the tumor cells were detected by conducting *in vitro* experiments.

Results: The *E. coli* infection score was lower in tumor samples than in normal samples. Eight hub genes were selected from a total of 28 genes associated with *E. coli*-related dysbiosis in gastrointestinal tumors to establish an accurate diagnostic model. The AUC values of *PRKCB* and *IL16* were all greater than 0.7 in three external datasets and the mRNA expression pattern was consistent with TCGA cohort, therefore *PRKCB* and *IL16* were selected as the diagnostic biomarkers. *PRKCB* and *IL16* exhibited significant positive correlations with most immune cells, and inflammation-related pathways were activated in the high expression groups of *PRKCB* and *IL16*. Moreover, *IL16* was high-expressed but *PRKCB* was low-expressed in STAD cells, and silencing *IL16* suppressed the invasion and migration of STAD cells.

Conclusions: Overall, we identified and validated 8 robust genes related to *E. coli* applying bioinformatics and machine learning algorithms, providing theoretical foundations for the relationship between *E. coli*-related dysbiosis and gastrointestinal tumors.

1. Introduction

Gastrointestinal tumors are the most common tumors with high morbidity and mortality rates worldwide [1,2]. According to the

* Corresponding author.
E-mail address: 1040175881@qq.com (L. Shi).

https://doi.org/10.1016/j.heliyon.2024.e40491
Received 25 June 2024; Received in revised form 13 November 2024; Accepted 15 November 2024
Available online 16 November 2024

global cancer statistics for 2020, colorectal cancer (CRC) and stomach adenocarcinoma (STAD) showed the highest incidence among all gastrointestinal tumors worldwide [3,4]. The occurrence of gastrointestinal tumors are related to multiple risk factors, such as smoking, alcohol consumption, high saturated fat diets, *Helicobacter pylori* infection, and inflammatory bowel disease [5,6]. At present, the relationship between gastrointestinal microbiome and tumors has gradually attracted increasing research attention [7]. The human gut is the home to roughly 30 trillion microorganisms [8]. Alterations in the gut microbiota and its byproducts play a crucial role in the onset of various illnesses, including carcinomas, gastrointestinal issues, metabolic disorders, and neurodegenerative diseases [9]. Growing studies indicated a notable connection between the development of CRC and gut microbiome maintained through several mechanisms, including DNA damage, inflammation, and related molecular pathways [10,11]. A specific strain of *Escherichia coli* (*E. coli*) contains a polyketide synthetase island that spans 54 kilobases. The island is responsible for the production of enzymes related to colibactin, a genotoxin composed of polyketide and peptide components [12,13]. *E. coli* is a major bacterial agent leading to foodborne illnesses worldwide [14]. Different strains of pathogenic *E. coli* exhibit a range of virulence factors and genes that allow them to induce diseases in humans, for instance, adhesion factors such as type 1 pili, P fimbriae, and S fimbriae enhance the colonization of epithelial cells [15]. Hence, a better understanding of the mechanism of *E. coli* during the progression of gastrointestinal tumors can improve the prevention and treatment for gastrointestinal tumors.

Weighted gene co-expression network analysis (WGCNA) and machine learning have been increasingly used to discover potential biomarkers for cancers [16,17]. WGCNA sections critical modules and identifies disease-relevant genes by establishing scale-free networks through connecting gene expressions with clinical features [18]. As a branch of artificial intelligence, machine learning focuses on using algorithms to create decision models to complete tasks [19,20]. In recent years, various machine learning algorithms such as randomForest and LASSO regression analysis have been extensively applied in screening biomarkers and constructing diagnostic models for malignant tumors [21,22]. For example, combining WGCNA with machine learning methods, Lv et al. found 6 biomarkers for the prediction, prevention, and treatment of liver hepatocellular carcinoma (LIHC) [23]. However, identification of biomarkers related to *E. coli* infection for diagnosing gastrointestinal tumors based on biological network analysis and machine learning algorithms has not been reported.

Alterations in complex systems, including the relationship between gut microbiota and its host, help understand the fundamental molecular processes. In gastrointestinal diseases, the composition of gut microbiota experiences shifts at the phylum level, a phenomenon known as dysbiosis [24]. Therefore, this study set out to discover new diagnostic biomarkers associated with *E. coli*-related dysbiosis in gastrointestinal tumors. Firstly, the co-expressed gene modules were classified to identify module genes showing the highest correlation with the diseases. By intersecting the differentially expressed genes (DEGs) with the module genes, the DEGs associated with *E. coli*-related dysbiosis in gastrointestinal tumors were obtained. Then, hub genes were selected to establish a diagnostic model, which was further validated in external datasets, and the diagnosis biomarkers were determined. Additionally, we examined the correlation between the biomarkers and immune cells in tumor samples to analyze their potential role in the immune microenvironment of gastrointestinal tumors and possible immune regulatory mechanisms. Finally, the mRNA expressions of biomarkers in STAD cells were validated and the tumor cell invasion and migration were assessed by conducting *in vitro* assays.

2. Material and methods

2.1. Data collection and data preprocessing

The standardized pan-cancer datasets of The Cancer Genome Atlas (TCGA) Pan-Cancer were downloaded from the University of California Santa Cruz (UCSC) database (https://xenabrowser.net/). The TCGA Pan-Cancer dataset contained 1622 tumor samples and 144 normal samples. The Fragments Per Kilobase per Million (FPKM) sample data of Solid Tissue Normal and Primary Tumor of TCGA-ESCA, TCGA-COAD, TCGA-STAD, TCGA-READ, TCGA-PAAD, and TCGA-LIHC were retained. Subsequently, after removing samples without clinical follow-up data, those with a survival time longer than 0 were retained. Then, the TCGA datasets were merged and the batch effect was removed using the "removeBatchEffect" function of the Limma package (Fig. S1) [25].

GSE66229 dataset (100 normal samples and 300 tumor samples), GSE44076 dataset (148 normal samples and 98 tumor samples), and GSE54236 dataset (80 normal samples and 81 tumor samples) were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) [26]. For GEO datasets, the annotation information of the corresponding chip platform was downloaded. The probes were mapped to genes based on the annotation information while removing the probes matched to multiple genes. The average value of the multiple probes that matched one gene was calculated as the expression value of the gene. It should be noted that some datasets used in this study had a small sample size that may amplify individual differences and cause potential bias. For this reason, we are planning to include more cohorts for validation in our future studies to promote the generalizability of the current findings.

2.2. Calculation of E. coli score, StromalScore, ImmuneScore, ESTIMATEScore, and TumorPurity

The *E. coli* infection gene set with 56 genes was downloaded from Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.ad.jp/kegg/). The *E. coli* score for all samples in TCGA cohort was calculated by ssGSEA [27]. Meanwhile, the StromalScore, ImmuneScore, and ESTIMATEscore were computed by performing ssGSEA using ESTIMATE algorithm in the 'estimated' R package [28], the TumorPurity was calculated based on the ESTIMATEScore [29]. Finally, the correlation analysis between these scores in tumor samples was conducted.

2.3. Development of a gene co-expression network

WGCNA package in R [30] was employed to section co-expressed gene modules of tumor samples. Spearman method was used to analyze the module-trait relationships between the modules and *E. coli* score and the correlation coefficients were visualized by plotting a heatmap. Next, module-trait associations were determined based on the correlation between gene significance (GS) for *E. coli* score and module membership (MM) in the critical module with the highest correlation coefficient. Finally, gene ontology (GO) enrichment analysis in biological process (BP) term was performed using the genes in the critical module with the clusterProfiler package in R software [31].

2.4. Identification of differentially expressed genes (DEGs)

Based on the mRNA expression profile data of TCGA, the DEGs analysis between tumor and normal samples were performed using Limma package in R software [32] under the screening threshold of p.adj<0.05 and $|\log2\text{FoldChange (FC)}|>1$. Then, the expressions of the top 50 DEGs were visualized by generating a heatmap [33]. Simultaneously, the DEGs between the two types of samples were intersected with the critical module genes of $|\text{MM}|\geq0.8$ to obtain the DEGs associated with *E. coli*-related dysbiosis in gastrointestinal tumors.

2.5. Screening key markers through machine learning methods

Random forest is an integrated learning algorithm capable of processing high dimensional datasets without overfitting problems. It improves the accuracy and stability of a model by creating multiple decision trees and taking their votes [34]. LASSO regression can automatically reduce the regression coefficients of irrelevant or less relevant genes to zero for feature selection. This is particularly important when processing biological data containing a large number of features, and it can effectively filter out the most valuable biomarkers [35]. Hence, we combined these two methods to identify key genes with diagnostic and prognostic value. Firstly, a randomForest model was established using randomForest function in the R software package randomForest [36]. The number of variables (mtry) used for binary trees in the specified nodes of the model and decision trees (ntree) included in the randomForest were determined. The mtry was ascertained according to the average misjudgment rate of the model based on Outofbag (OOB) samples, and the ntree was determined according to the correlation between the number of decision trees and model error. Next, the randomForest model was built with the determined variables. The top 10 markers were selected based on MeandecreaseAccuracy and MeandecreasGini, respectively. Then, the key disease markers were screened by LASSO regression analysis. The feature factors were identified with 10-fold cross validation using R software package glmnet (parameters: nfolds = 10 and family = 'binary') [37]. The number of key disease markers was determined according to the λ value corresponding to the minimum binomial deviance.

2.6. Establishment and validation of diagnostic model

The candidate hub genes with diagnostic value were identified from the intersection of the markers obtained by randomForest and LASSO regression methods. Then, receiver operating characteristic (ROC) analysis was applied to assess the diagnosis accuracy of each hub gene and the model. The area under the ROC curve (AUC) was calculated, with an AUC value > 0.7 indicating a high prediction accuracy of the model [38]. The differential expressions of the hub genes between normal and tumor samples were displayed by boxplot. Moreover, the diagnostic value of the hub genes was validated in the external datasets of GSE66229, GSE44076, and GSE54236.

2.7. GSEA and immune infiltration analysis

Spearman method [39] was employed to examine the correlation between the identified biomarkers and immune cells in tumor samples. Based on the median gene expression of the biomarkers, the TCGA samples were divided into low and high expression groups. Next, GSEA was used for biological function and pathway enrichment analysis in the background gene set (FDR<0.05). The "h.all.v2023.1.Hs.entrez.gmt" was a reference gene set obtained from the Molecular Signature Database (MsigDB, https://www.gsea-msigdb.org/gsea/msigdb/) [40].

2.8. Cell culture and transfection

DMEM medium comprising 10 % fetal bovine serum (CBP60512M) was used to culture normal human gastric epithelial cell line GES-1 (CBP60512), whereas STAD cell line AGS (CBP60476) were cultured in RPMI-1640 medium comprising 10 % fetal bovine serum (CBP60476M). The two types of cells were commercially purchased from Nanjing Cobioer Biosciences Co., Ltd. (Nanjing, China). All the cells were incubated at 37 °C in an incubator in 5 % CO_2.

Next, cell transfection was performed for silencing *IL16*. The small interfering (si) RNA targeting *IL16* (si-*IL16*, target sequence: 5′-GAGGAGTTACTTTGTAGTTTTAA-3′) and negative control (si-NC) were synthesized by Sangon Biotech (Shanghai) Co., Ltd. The cells were transfected for 48 h (h) with the use of Lipofectamine 2000 transfection assay kit (Invitrogen, Thermo Fisher Scientific) following the description.

2.9. Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

The total RNAs of GES-1 and AGS cells were extracted by TriZol reagent (15596018CN, Invitrogen) and converted into cDNA applying Maxima First Strand cDNA Synthesis Kit (K1641, Invitrogen). Further, 1 μL cDNA was applied for qRT-PCR with SYBR Green Supermix (11744500, Invitrogen) following the specifications. The parameters of qRT-PCR were as follows: denaturation at 95 °C for 5 min (min), 40 cycles of 95 °C for 15 s (s), and annealing at 60 °C for 30s, then elongation at 75 °C for 30s. The primer sequences were presented in Supplementary Table 1. The relative mRNA expressions of *IL16* and *PRKCB* were quantified with the $2^{-\Delta\Delta CT}$ method, with β-actin as a normalization control [41].

2.10. Transwell assay

Transwell assay was performed to detect the invasion of AGS cells [42]. First, 80 μL Matrigel (diluted at 1:8) was slowly added into the Transwell chamber (with 8 μm filters) and incubated at 37 °C for 1h to solidify the Matrigel. Then, the transfected AGS cell (1×10^5 cells/well) was transferred to the upper chamber with 200 μL serum-free medium, and 700 μL RPMI-1640 medium containing 10 % fetal bovine serum was added in the lower chamber. After culturing the cells for 24h, the AGS cells in the lower chamber were fixed with 5 % formaldehyde for 30min and stained with 1 % crystal violet for 15min. Finally, 10 regions were randomly selected to observe the numbers of AGS cells in the lower chamber under an Axiovert 5 digital microscope (ZEISS, Oberkochen, Germany).

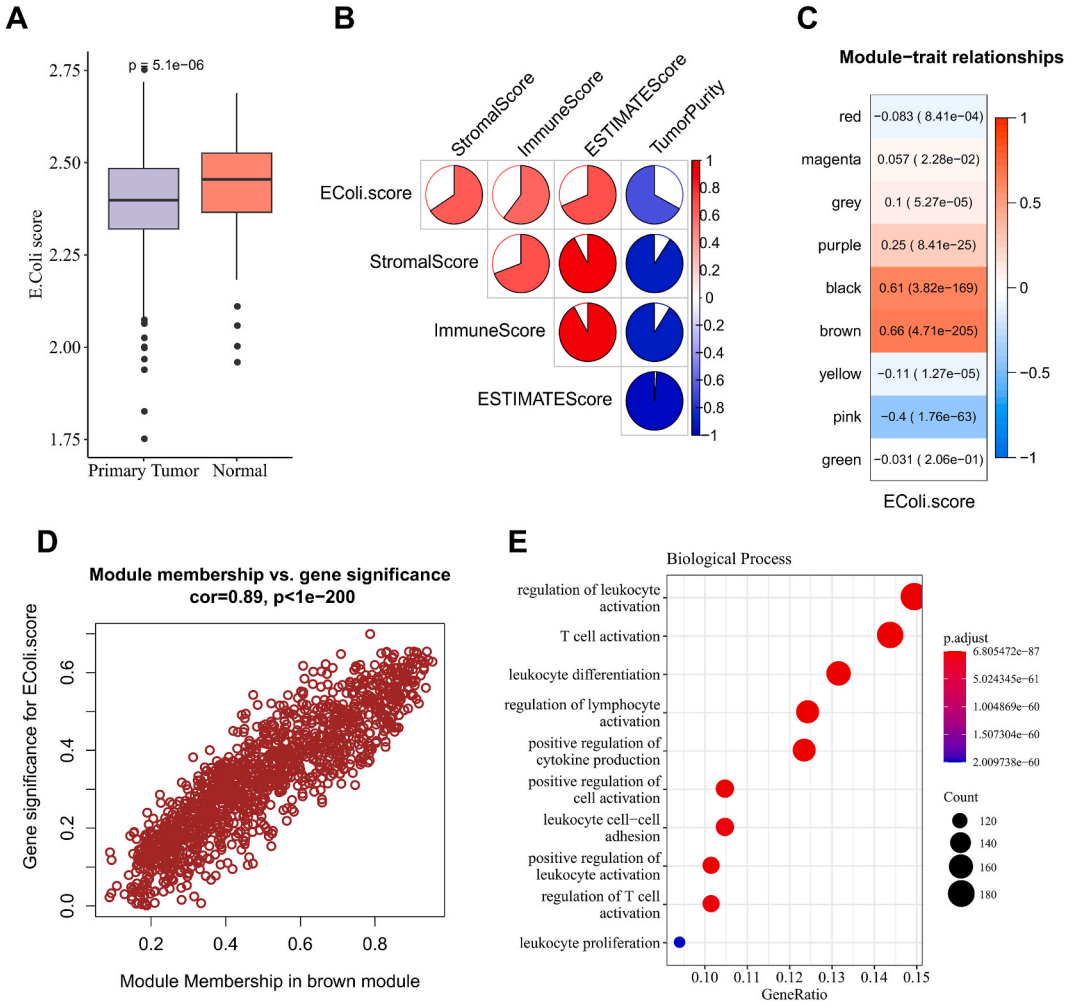


Fig. 1. Construction of co-expressed gene modules through WGCNA analysis. (A) The difference in *E. coli* score between tumor and normal samples. (B) The correlation of *E. coli* score, StromalScore, ImmuneScore, ESTIMATEScore, and TumorPurity in tumor samples. (C) The relationship between different modules and *E. coli* score. (D) Scatter diagram of the module membership (MM) in brown module versus gene significance for *E. coli* infection score. (E) GO enrichment results of brown module in the biological process. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.11. Wound healing assay

The migration of AGS cells was evaluated by wound-healing assay [43]. In brief, the AGS cells were inoculated to a 6-well plate with 800 μL cell suspension in each well and cultured for 24h. Then, the cells were scratched to create a wound using an aseptic 100 μL pipette and then maintained in serum-free medium at 37 °C. Next, the representative images were captured at 0 and 48h using the Axiovert 5 digital microscope (ZEISS, Oberkochen, Germany) and the wound closure (%) of the AGS cells were calculated.

2.12. Statistical analysis

R software (version 3.6.0) was employed in bioinformatics analyses. Differences between two groups of continuous variables were calculated using Wilcoxon rank-sum test. All the data were expressed as mean ± standard deviation (SD), and statistical analyses were conducted using SPSS (version, SPSS, Inc., Chicago, IL). Two-group differences were compared by student's t-test. $P < 0.05$ was considered as statistically significant.

3. Results

3.1. Gene modules closely correlated with E. coli infection in gastrointestinal tumors

Firstly, the *E. coli* infection score of all TCGA samples was calculated using ssGSEA method. Comparison on the difference of *E. coli* score between primary tumor samples and normal samples showed that tumor samples had lower *E. coli* infection score compared to normal samples (Fig. 1A). Specifically, the *E. coli* infection score was strongly positively related to StromalScore, ImmuneScore, and ESTIMATEScore but negatively related to TumorPurity in tumor samples (Fig. 1B). Next, a total of 9 co-expressed gene modules in tumor samples was classified by WGCNA method, with the brown module exhibiting the highest correlation (r = 0.66) with *E. coli* score (Fig. 1C). Thus, the brown module was selected for subsequent research. Next, the correlation analysis demonstrated a significant positive correlation between the GS for *E. coli* score and MM in the brown module (correlation coefficient = 0.89, Fig. 1D). These results suggested that the genes in the brown module were most closely associated with *E. coli*-related dysbiosis in gastrointestinal tumors. Furthermore, GO enrichment analysis in BP term was also performed based on the genes in brown module. The results showed that the pathways related to the proliferation and activation of immune cells such as regulation of leucocyte activation and T cells activation were significantly enriched (Fig. 1E).

3.2. The DEGs associated with E. coli-related dysbiosis in gastrointestinal tumors were identified

A total of 3995 DEGs between tumor and normal samples from the TCGA dataset were filtered by the Limma package for subsequent analysis (Fig. 2A). The expressions of the top 50 DEGs were shown in a heatmap (Fig. 2B). Subsequently, 28 DEGs associated with *E. coli*-related dysbiosis in gastrointestinal tumors were selected by intersecting the 3995 DEGs with the genes of $|MM| \geq 0.8$ in the brown module (Fig. 2C).

3.3. Key genes for the diagnosis of gastrointestinal tumors were screened by machine learning methods

Critical genes for diagnosing gastrointestinal tumors were determined applying machine learning algorithms of randomForest and LASSO regression. As shown in Fig. 3A, a random forest model was constructed with the 28 DEGs utilizing the randomForest function. Subsequently, the top 10 marker genes with the greatest contribution were selected based on MeandecreaseAccuracy and MeandecreaseGini. In addition, LASSO regression analysis was also used to identify the disease-related genes. We found that the binomial deviance was the smallest when the λ was 19, therefore the number of key markers was determined to be 19 (Fig. 3B).

3.4. Establishment of a diagnostic model using the hub genes for detecting gastrointestinal tumors

A total of 8 hub genes (including *MS4A6A*, *PRKCB*, *CD4*, *SCIMP*, *IL16*, *MPEG1*, *CD48,* and *LY9*) were collected by intersecting the top 10 most important genes from random forest with the 19 genes from LASSO regression (Fig. 4A). Then, the ROC curve was used to reflect the diagnostic performance of each hub gene and the diagnostic model, with an AUC value of 0.84, 0.78, 0.77, 0.77, 0.75, 0.73, 0.73, and 0.71 for *PRKCB*, *IL16*, *MPEG1*, *MS4A6A*, *CD4*, *SCIMP*, *LY9*, and *CD48*, respectively (Fig. 4B). These results indicated that the 8 hub genes and the model had a high diagnostic accuracy. Further analysis on the mRNA expressions of the 8 hub genes in the tumor and normal groups showed that the relative mRNA expressions of these genes were lower in the tumor group compared to the normal group (Fig. 4C).

3.5. The diagnostic model was validated and 2 diagnosis biomarkers for gastrointestinal tumors were identified

Furthermore, the diagnostic performance of the 8 hub genes was validated in the external datasets of GSE66229, GSE44076, and GSE54236. It was found that only the AUC values of *PRKCB* and *IL16* were greater than 0.7 in the three datasets (Fig. 5A–C, and E), and that the mRNA expression trend of the two genes was consistent with the data in the TCGA cohort (Fig. 5B–D, and F). Therefore, *PRKCB* and *IL16* could be considered as the biomarkers for the diagnosis of gastrointestinal tumors.

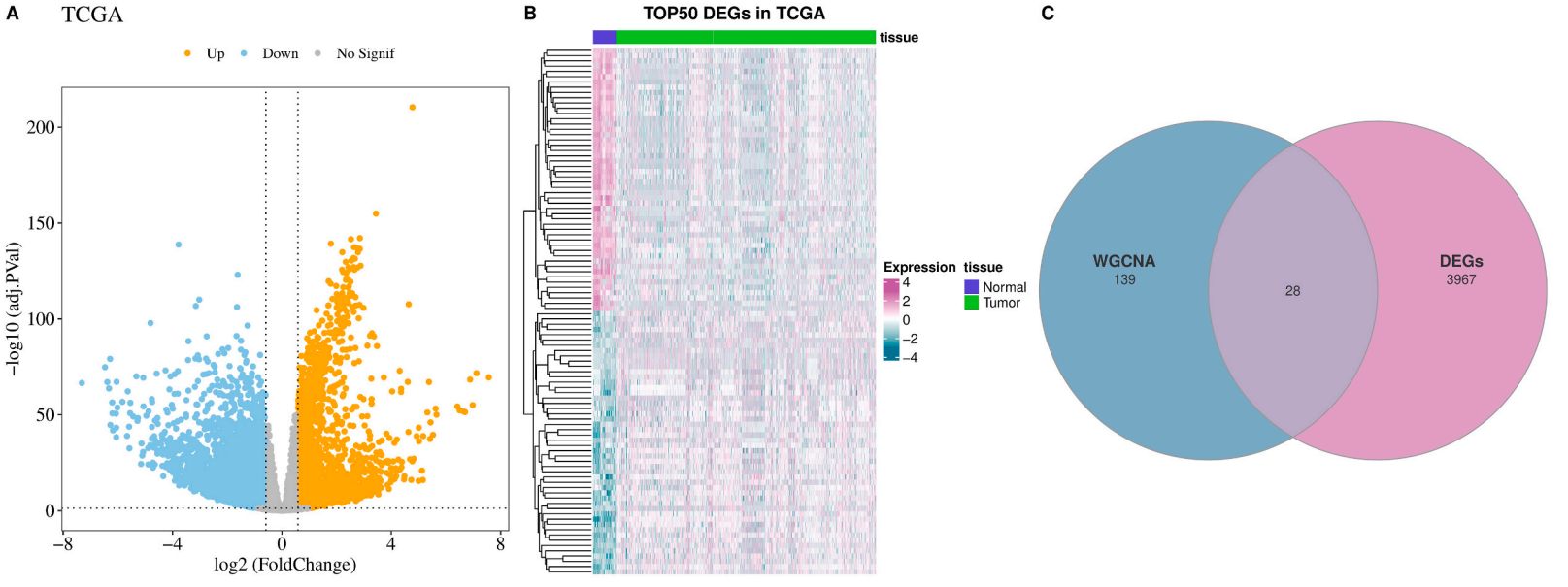


Fig. 2. DEGs analysis in the tumor and normal groups. (A) Volcano plot of DEGs distribution in TCGA cohort. (B) Heatmap of expression levels of top 50 DEGs. (C) Overlapping genes in DEGs and WGCNA with $|MM| \geq 0.8$.

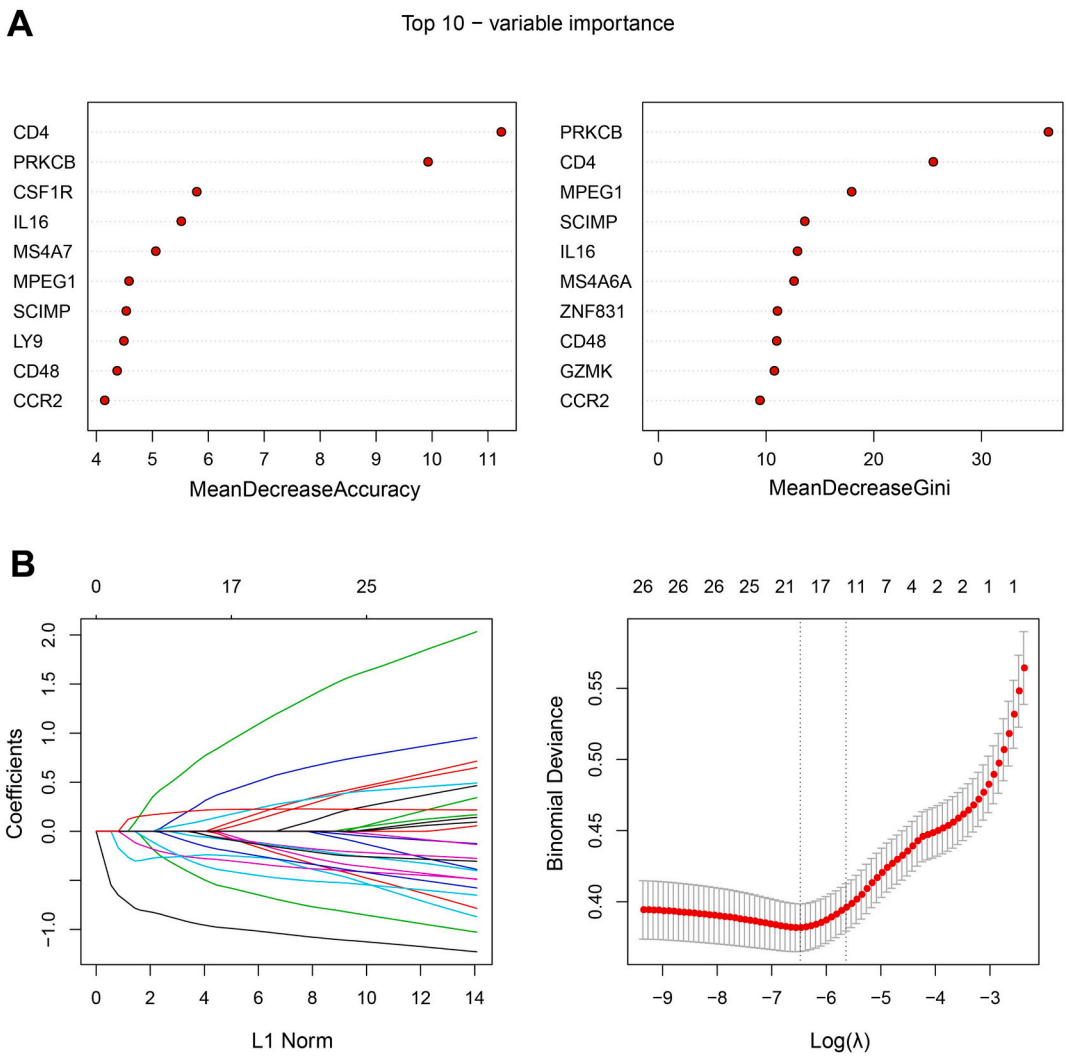


Fig. 3. Machine learning to screen key diagnosis genes for gastrointestinal tumors. (A) The top 10 contribution markers selected by randomForest method based on MeandecreaseAccuracy and MeandecreaseGini, respectively. (B) Coefficient profiles of different genes varying with λ and fitting curve of LASSO regression model coefficients.

3.6. Correlation between the pathogenic E. coli infection-related markers and the immune microenvironment

The correlation between the 2 biomarker genes (*PRKCB*, *IL16*) and 28 immune cells in tumor samples was also analyzed applying the ssGSEA method. It was found that *PRKCB* and *IL16* showed a significant positive correlation with most immune cells, such as Regulatory T cells, Activated CD4 T cells, Macrophages, Natural killer cells, Activated CD8 T cells, Mast cells, T follicular helper cells, and Activated B cells, (Fig. 6A). The samples were divided into high and low expression groups of *PRKCB* and *IL16* based on the median gene expression to further compare the differences in biological functions and signaling pathways involved in the different expression groups in TCGA dataset. Next, GSEA was used for enrichment analysis (FDR<0.05). It was found that the inflammation-related pathways such as INTERFERON_GAMMA_RESPONSE, INFLAMMATORY_RESPONSE, and IL6_JAK_STAT3_SIGNALING were activated in the high expression groups of *PRKCB* and *IL16* (Fig. 6B and C).

3.7. Silencing IL16 inhibited STAD cell invasion and migration

The results of qRT-PCR experiment demonstrated that *IL16* was expressed significantly but *PRKCB* was expressed notably lower in STAD cells AGS (Fig. 7A and B). Furthermore, the Transwell and wound healing assays showed that silencing *IL16* reduced the numbers of invaded and migrated AGS cells (Fig. 7C–F).

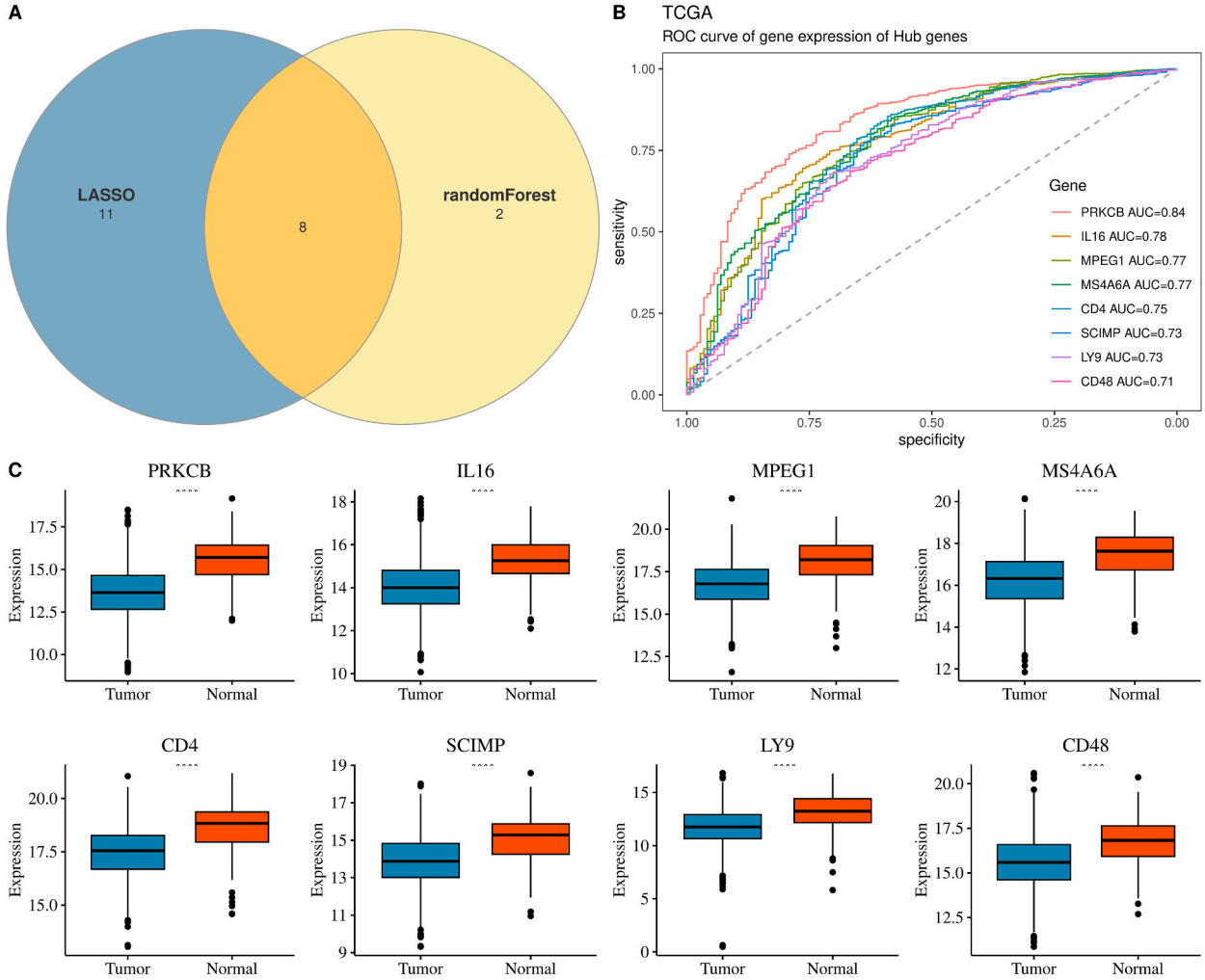


Fig. 4. Establishment of diagnostic model. (A) Venn diagram of the intersection of diagnostic markers obtained by randomForest and LASSO regression methods. (B) ROC curve of the 8 hub genes. (C) Boxplot of mRNA expression levels of the 8 hub genes.

4. Discussion

Gastrointestinal tumors have a high incidence rate [44]. Gastric cancer patients exhibit different microbiota compositions (ecological dysbiosis) than healthy population [45]. Variations in the makeup of gut microbial communities together with the bacterial extracellular vesicles they release [46] interact closely with tumor microenvironment, directly leading to the reprogramming of tumor microenvironment and profoundly influencing cancer immunity of gastric cancer [47]. Although gut bacteria play a critical role in regulating the intestinal microenvironment and the health of gastric cancer patients, specific mechanisms of such a regulation remained largely unexplored. Therefore, this study identified the key genes associated with the gastrointestinal tumor pathogenic *E. coli*-related dysbiosis applying bioinformatics analyses and machine learning algorithms and then established a diagnostic model with a strong diagnostic performance. In addition, we also analyzed the immune cells to explore the potential mechanisms of pathogenic *E. coli*-related dysbiosis that regulated the development and progression of gastric cancer.

Machine learning algorithms have been widely applied in screening biomarkers for the diagnosis and prognostic prediction of patients with gastrointestinal malignant tumors [48]. Previous research by Gilani identified potential miRNA biomarkers for STAD diagnosis and confirmed hsa-miR-1343-3p as a biomarker with a high precision for detecting STAD [49]. Zheng et al. [50] constructed a prognosis signature consisting of 6 long non-coding RNAs (lncRNAs, LINC01273, AL445524.1, PSPC1-AS2, AC109439.2, LINC00547, and AC015922.3), which was significantly correlated with the prognosis of ESCA and had the potential to be an supplementary biomarker for Tumor-Node-Metastasi (TNM) stage to subdivide ESCA patients more accurately. Additionally, based on WGCNA and machine learning, Zheng et al. discovered 5 immune-related molecular biomarkers (*EFHD1*, *KIF4A*, *UBE2C*, *SMYD3*, and *MCM7*) for improving the diagnosis, prognosis, and molecular therapy of LIHC [51]. Importantly, studies have shown a complex interaction between gastrointestinal tumors and gut microbiota [52,53], but using genes related to gut microbiota as biomarkers for

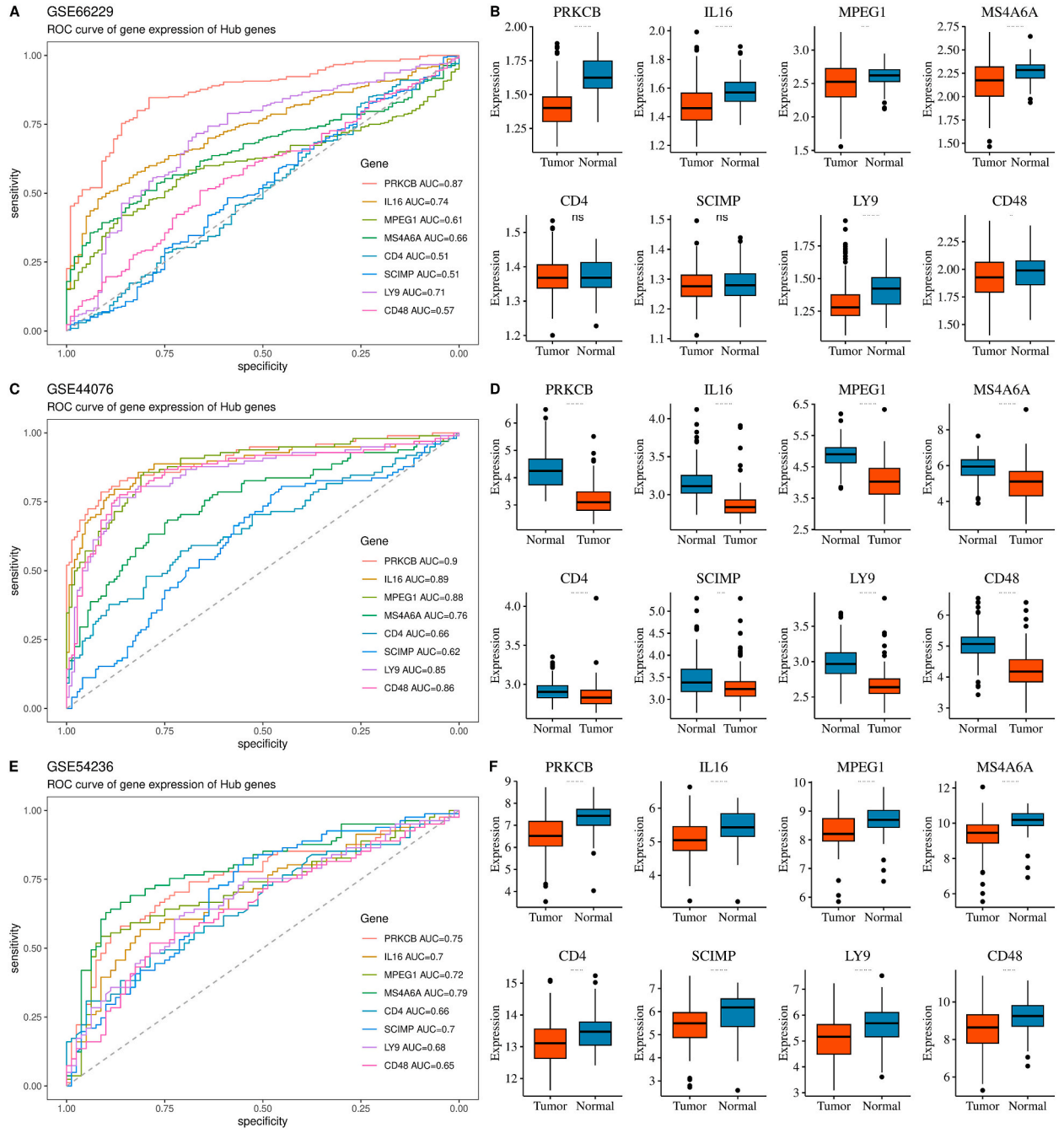


Fig. 5. Diagnostic model validation in external datasets. (A–B) ROC curve and expression levels of 8 hub genes in GSE66229. (C–D) ROC curve and expression levels of 8 hub genes in GSE44076. (E–F) ROC curve and expression levels of 8 hub genes in GSE54236. ns Means not significant.

the diagnosis and prognosis of gastrointestinal tumors was less explored. Recently, Liu et al. identified *IL7* and *BCL10* as the gut microbes-related biomarkers for CRC, and the two could predict the immunotherapeutic response and prognosis for CRC patients [54]. In our present study, Pan-Cancer datasets (TCGA-ESCA, TCGA-READ, TCGA-PAAD TCGA-LIHC, TCGA-STAD, and TCGA-COAD) containing 144 normal samples and 1622 tumor samples were downloaded. We identified 8 hub genes (*PRKCB*, *IL16*, *MPEG1*, *MS4A6A*, *CD4*, *SCIMP*, *LY9*, *CD48*) with low expression levels in tumor group and established a diagnostic model applying WGCNA and machine learning methods. After verifying the diagnostic performance of these genes in three external datasets, *PRKCB* and *IL16* were selected as the biomarkers related to *E. coli* infection for the detection of gastrointestinal tumors.

As a member of the protein kinase family, protein kinase C-B (*PRKCB*) consists of several serine- and threonine-kinases and can be activated by calcium and the second messenger diacylglycerol [55,56]. Previous research [57] showed that highly methylated and

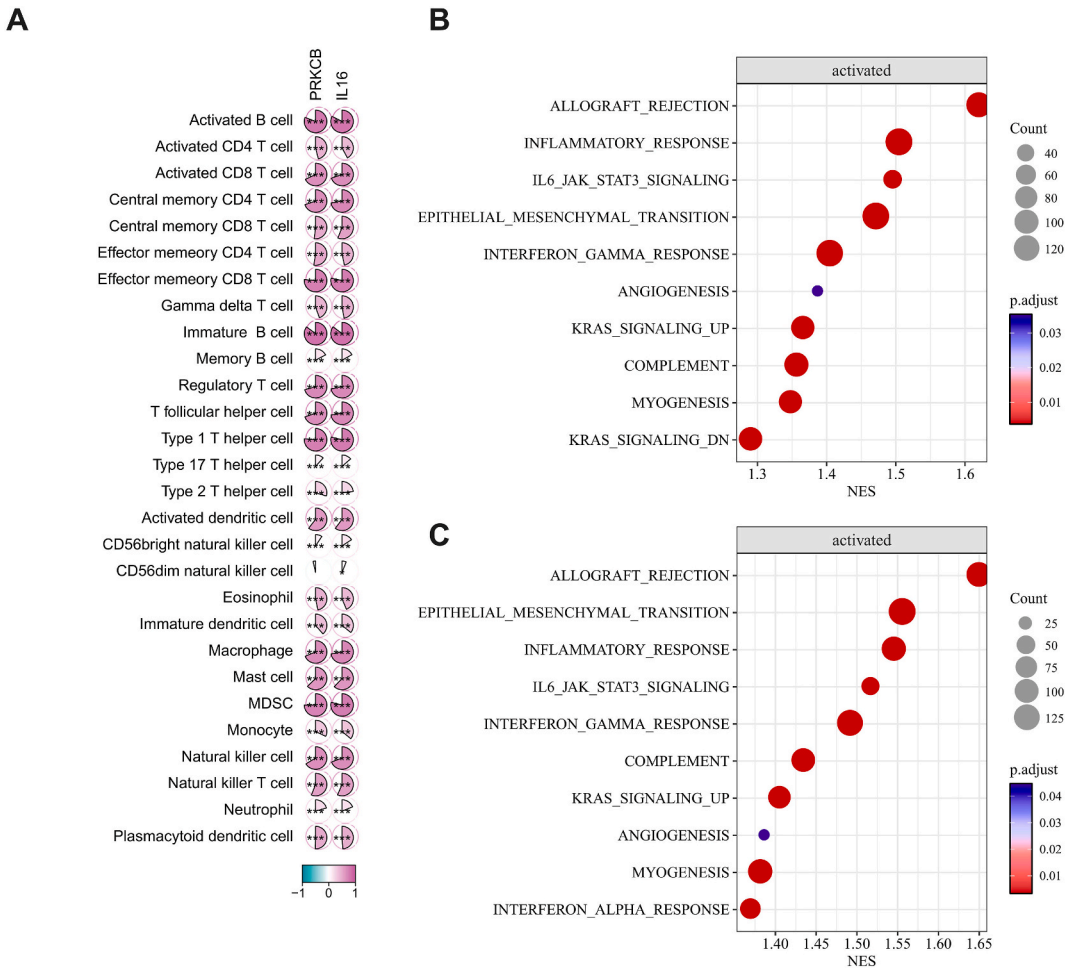


Fig. 6. GSEA and immune analysis in tumor samples. (A) Correlation between biomarkers and 28 immune cell scores. *** Means $p < 0.001$; * Means $p < 0.05$. (B–C) PRKCB (B) and IL16 (C) genes were categorized into high and low expression groups, respectively, based on the median values of PRKCB and IL16 gene expression in the TCGA dataset for GSEA pathway analysis.

low-expressed *PRKCB* in lung adenocarcinoma (LUAD) was associated with immune cell infiltration, and that LUAD patients with a low *PRKCB* expression tended to have a worse survival prognosis. Fan et al. reported lower mRNA and protein levels of *PRKCB* in CRC tissues than in normal tissues [58]. Consistently, the present study revealed that *PRKCB* had a lower mRNA expression in tumor group than in normal group. Such a finding was further verified using STAD cells by performing qRT-PCR, the data of which supported that *PRKCB* was significantly positively correlated with most immune cells. The low mRNA expression of *PRKCB* was caused by its high methylation that affected the normal expression and function of *PRKCB* in tumor cells, which might be involved in the occurrence and progression of tumors [59]. *IL16* is an immunomodulatory cytokine and a chemoattractant for $CD4^+$ immune cells [60]. *IL16* signaling pathway is a key signaling pathway involved in the interactions between tumor-associated neutrophils and other cells [61]. Jia et al. proposed that during influenza A virus (IAV) infection, deficiency in IL16 increased the response of T helper 1 cells and cytotoxic T lymphocytes in addition to promoting the maturation of dendritic cells. This finding contributed to the understanding of how the host regulated T cell immune responses during infection of the Influenza A virus [62]. In addition, Milke et al. reported that *IL16*, a chemokine regulated by the RNA-binding protein tristetraprolin, promoted the infiltration of monocytes and macrophages, which in turn has been associated with a poor prognosis of breast cancer patients [63]. Similarly, we observed that *IL16* expression in gastric cancer was significantly and positively correlated with the degree of immune infiltration of a variety of cells including monocytes, T cells, macrophages, and B cells. This suggested that *IL16* played an important role in the progression of *E. coli*-triggered gastric cancer by modulating immune responses. Furthermore, in this study, the mRNA expression of *IL16* was lower in tumor group than the normal group in both TCGA and GEO datasets. However, our *in vitro* validation results showed that the mRNA expression level of *IL16* was upregulated in STAD cells, and that silencing *IL16* suppressed the invasion and migration of STAD cells. According to Fang et al. [64], the protein level of *IL16* was significantly higher in STAD tissues than in adjacent normal tissues (n = 7). Fu et al. also indicated that compared to healthy control, elevated serum levels of *IL16* were significantly related to tumor recurrence and poor prognosis of STAD patients [65]. These results indirectly confirmed that *PRKCB* and *IL16* could influence the *E. coli*-associated tumor microenvironment

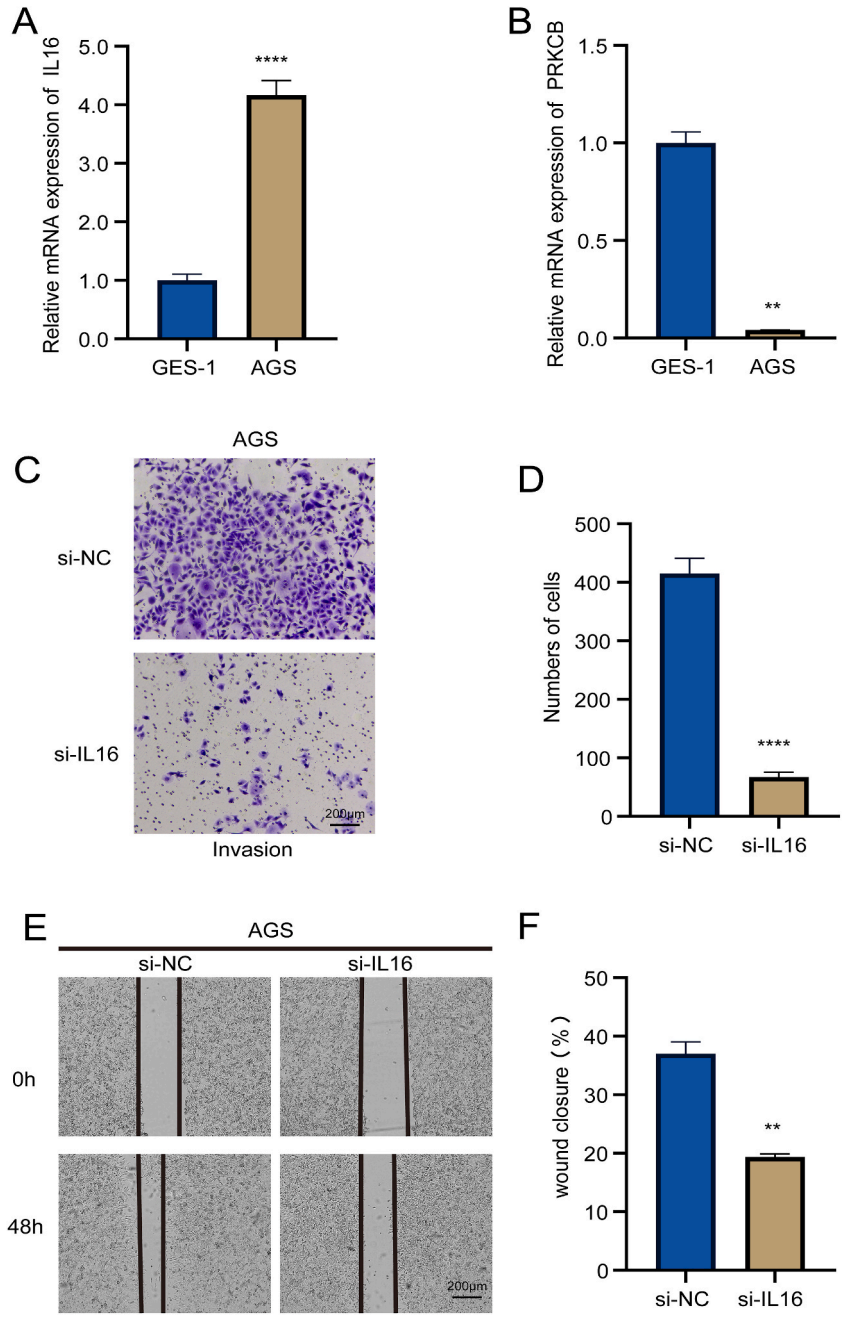


Fig. 7. *In vitro* validation in STAD cells. (A–B) Relative mRNA expression levels of 2 biomarkers, *IL16* (A) and *PRKCB* (B), in STAD cells AGS and normal human gastric epithelial cells GES-1 via qRT-PCR. (C–D) Effects of *IL16* silencing on the invasion of STAD cells AGS assessed by Transwell assay. (E–F) Effects of *IL16* silencing on the migration of STAD cells AGS detected by wound healing assay. **** Means $p < 0.0001$; ** Means $p < 0.01$.

of gastrointestinal tumors by regulating immune cell infiltration and immune responses, which in turn influenced tumor progression.

Our study also had some limitations. Although datasets from the TCGA and GEO databases were used for analysis and validation, a small sample size of some datasets may not be able to fully reflect the variability of gene expressions in different populations. In particular, the small sample size of certain subtypes of gastrointestinal tumors could limit the broad reliability and applicability of the current results. Therefore, future studies should expand the sample size by recruiting patients of different races, genders, ages, and tumor subtypes. In addition, the significance of the present findings obtained at the transcriptome level required further validation from clinical and basic experiments. Importantly, the datasets analyzed in this study was primarily tumor tissue samples from patients, whereas the cell line experiments were performed in an *in vitro* setting, and such a setting difference may lead to inconsistencies in IL6

expression levels. In the future, different levels of histological data such as proteomics, epigenetics, and metabolomics can be integrated in order to more comprehensively explore the expression differences of IL6 in tumors. This may help to reveal the expression patterns of IL6 at different regulatory levels and explore the specific mechanisms of its role in the tumor microenvironment.

5. Conclusion

In conclusion, this study identified 8 hub genes (including *PRKCB*, *IL16*, *MPEG1*, *MS4A6A*, *CD4*, *SCIMP*, *LY9*, and *CD48*) related to *E. coli* infection in gastrointestinal tumor to establish a diagnostic model, and the diagnostic performance of the model was validated in external datasets. Importantly, *PRKCB* and *IL16* played important roles in immunomodulation and tumor progression. The present study provided new insights into the relationship between *E. coli* infection and gastrointestinal tumors and emphasized the potential roles of key genes in the immune microenvironment, however, these findings should be further explored in depth to ensure their broad applicability and clinical translational potential.

CRediT authorship contribution statement

Tingting Ge: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Methodology, Data curation, Conceptualization. **Wei Wang:** Validation, Supervision, Resources, Methodology, Investigation. **Dandan Zhang:** Visualization, Supervision, Resources, Methodology. **Xubo Le:** Visualization, Software, Methodology, Formal analysis, Conceptualization. **Lumei Shi:** Writing – original draft, Validation, Resources, Investigation, Formal analysis.

Ethics statement

None.

Data availability statement

The datasets generated during and/or analyzed during the current study are available in the GSE repository [GSE66229] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE66229) and [GSE44076] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE44076).

Funding

The author declare that they have received no funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e40491.

Abbreviations

AUC	area under the ROC curve
COAD	colon cancer
DEGs	differentially expressed genes
ESCA	esophageal cancer
ESTIMATE	Estimation of STromal and Immune cells in MAlignant Tumors using Expression data
FC	FoldChange
FPKM	Fragments Per Kilobase per Million
GEO	Gene Expression Omnibus
GO	gene ontology
GS	gene significance
GSEA	Gene Set Enrichment Analysis
lncRNA	long non-coding RNAs

KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least Absolute Shrinkage and Selection Operator
LIHC	liver cancer
LUAD	lung adenocarcinoma
MM	module membership
NAFLD	non-alcoholic fatty liver disease
OOB	Outofbag
PAAD	pancreatic cancer
PRKCB	Protein kinase C-B
qRT-PCR	quantitative real-time polymerase chain reaction
READ	rectal cancer
ROC	receiver operating characteristic
SD	standard deviation
si	small interfering
ssGSEA	single sample Gene Set Enrichment Analysis
STAD	stomach adenocarcinoma
TCGA	the Cancer Genome Atlas
TNM	Tumor-Node-Metastasi
UCSC	University of California Santa Cruz
WGCNA	weighted gene co-expression network analysis

References

[1] C.M. Ulrich, C. Himbert, A.N. Holowatyj, S.D. Hursting, Energy balance and gastrointestinal cancer: risk, interventions, outcomes and mechanisms, Nat. Rev. Gastroenterol. Hepatol. 15 (11) (2018) 683–698.

[2] M. Li, J. Wei, G. Xu, Y. Liu, J. Zhu, Surgery combined with molecular targeted therapy successfully treated giant esophageal gastrointestinal stromal tumor 24 (2) (2022) 349–356.

[3] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (3) (2021) 209–249.

[4] K. Liu, S. Song, T. Fu, Y. Liu, H. Zhang, M. Yan, et al., Spatiotemporal trends in the incidence of gastrointestinal neoplasms in wuwei city of northwestern China from 1995 to 2016: a hospital-based retrospective observational study, Front. Oncol. 11 (2021) 712857.

[5] S. Hanan, E. Luvsannyam, M.S. Jain, S. Laller, T. Cheema, C. Mellon, et al., A rare case of three distinct gastrointestinal neoplasms occurring simultaneously in an elderly patient, J. Med. Cases 12 (10) (2021) 419–423.

[6] A. Jafari-Sales, A. Shariat, H.-B. Baghi, B. Baradaran, B. Jafari, The presence of human papillomavirus and epstein-barr virus infection in gastric cancer, A Systematic Study 24 (3) (2022), 413–4126.

[7] V.T. Pham, S. Dold, A. Rehman, J.K. Bird, R.E. Steinert, Vitamins, the gut microbiome and gastrointestinal health in humans, Nutr. Res. (N.Y.) 95 (2021) 35–53.

[8] R. Sender, S. Fuchs, R. Milo, Revised estimates for the number of human and bacteria cells in the body, PLoS Biol. 14 (8) (2016) e1002533.

[9] Y. Chen, J. Zhou, L. Wang, Role and mechanism of gut microbiota in human disease, Front. Cell. Infect. Microbiol. 11 (2021) 625913.

[10] R.M. Bleich, J.C. Arthur, Revealing a microbial carcinogen, Science (New York, N.Y.) 363 (6428) (2019) 689–690.

[11] A.L.B. Bennedsen, S. Furbo, T. Bjarnsholt, H. Raskov, I. Gögenur, L. Kvich, The gut microbiota can orchestrate the signaling pathways in colorectal cancer, APMIS : acta pathologica, microbiologica, et immunologica Scandinavica 130 (3) (2022) 121–139.

[12] P.J. Dziubańska-Kusibab, H. Berger, F. Battistini, B.A.M. Bouwman, A. Iftekhar, R. Katainen, et al., Colibactin DNA-damage signature indicates mutational impact in colorectal cancer, Nat. Med. 26 (7) (2020) 1063–1069.

[13] T. Miyasaka, T. Yamada, K. Uehara, H. Sonoda, A. Matsuda, S. Shinji, et al., Pks-positive Escherichia coli in tumor tissue and surrounding normal mucosal tissue of colorectal cancer patients, Cancer Sci. 115 (4) (2024) 1184–1195.

[14] A.R. Manges, J.R. Johnson, Food-borne origins of Escherichia coli causing extraintestinal infections, Clin. Infect. Dis. : an official publication of the Infectious Diseases Society of America 55 (5) (2012) 712–719.

[15] P. Rathore, A. Basnet, A. Kilonzo-Nthenge, K. Dumenyo, Z. Yadegari, A. Taheri, Rapid detection of pathogenic E. coli based on CRISPR Cas system, Front. Microbiol. 15 (2024) 1423478.

[16] X. Jin, Z. Huang, P. Guo, R. Yuan, Screening of novel biomarkers for breast cancer based on WGCNA and multiple machine learning algorithms, Transl. Cancer Res. 12 (6) (2023) 1466–1489.

[17] Y. Fang, F. Xia, F. Tian, L. Qu, F. Yang, J. Fang, et al., Machine learning and bioinformatics to identify biomarkers in response to *Burkholderia pseudomallei* infection in mice 48 (4) (2024) 613–621.

[18] Y. Li, Y. Hu, F. Jiang, H. Chen, Y. Xue, Y. Yu, Combining WGCNA and machine learning to identify mechanisms and biomarkers of ischemic heart failure development after acute myocardial infarction, Heliyon 10 (5) (2024) e27165.

[19] C.H. Chang, C.H. Lin, H.Y. Lane, Machine learning and novel biomarkers for the diagnosis of alzheimer's disease, Int. J. Mol. Sci. 22 (5) (2021).

[20] X. Wang, H. Wang, S. Yu, X. Wang, Diagnosis model of paraquat poisoning based on machine learning, Curr. Pharmaceut. Anal. 18 (2) (2022) 176–181.

[21] M. Sherafatian, Tree-based machine learning algorithms identified minimal set of miRNA biomarkers for breast cancer diagnosis and molecular subtyping, Gene 677 (2018) 111–118.

[22] H. Wang, W. Cheng, P. Hu, T. Ling, C. Hu, Y. Chen, et al., Integrative analysis identifies oxidative stress biomarkers in non-alcoholic fatty liver disease via machine learning and weighted gene co-expression network analysis, Front. Immunol. 15 (2024) 1335112.

[23] J.H. Lv, A.J. Hou, S.H. Zhang, J.J. Dong, H.X. Kuang, L. Yang, et al., WGCNA combined with machine learning to find potential biomarkers of liver cancer, Medicine 102 (50) (2023) e36536.

[24] S.J. Taylor, M.G. Winter, C.C. Gillis, L.A.D. Silva, A.L. Dobbins, M.K. Muramatsu, et al., Colonocyte-derived lactate promotes E. coli fitness in the context of inflammation-associated gut microbiota dysbiosis, Microbiome 10 (1) (2022) 200.

[25] F. Fadil, C. Samol, R.S. Berger, F. Kellermeier, W. Gronwald, P.J. Oefner, et al., Isotope ratio outlier analysis (IROA) for HPLC-TOFMS-based metabolomics of human urine, Metabolites 12 (8) (2022).

[26] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, et al., NCBI GEO: archive for functional genomics data sets–update, Nucleic Acids Res. 41 (Database issue) (2013) D991–D995.

[27] B. Pan, Y. Li, Z. Xu, Y. Miao, H. Yin, Y. Kong, et al., Identifying a novel ferroptosis-related prognostic score for predicting prognosis in chronic lymphocytic leukemia, Front. Immunol. 13 (2022) 962000.
[28] X. Yuan, J. Zhou, L. Zhou, Z. Huang, W. Wang, J. Qiu, et al., Apoptosis-related gene-mediated cell death pattern induces immunosuppression and immunotherapy resistance in gastric cancer, Front. Genet. 13 (2022) 921163.
[29] K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.
[30] S. Jia, L. Zhai, F. Wu, W. Lv, X. Min, S. Zhang, et al., Integrative machine learning algorithms for developing a consensus RNA modification-based signature for guiding clinical decision-making in bladder, cancer 26 (2) (2024) 269–285.
[31] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, et al., clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innovation 2 (3) (2021) 100141.
[32] S. Guo, J. Wu, W. Zhou, X. Liu, Y. Liu, J. Zhang, et al., Identification and analysis of key genes associated with acute myocardial infarction by integrated bioinformatics methods, Medicine 100 (15) (2021) e25553.
[33] A. Zulibiya, J. Wen, H. Yu, X. Chen, L. Xu, X. Ma, et al., Single-cell RNA sequencing reveals potential for endothelial-to-mesenchymal transition in tetralogy of fallot 18 (6) (2023) 611–625.
[34] F.J. Brims, T.M. Meniawy, I. Duffus, D. de Fonseka, A. Segal, J. Creaney, et al., A novel clinical prediction model for prognosis in malignant pleural mesothelioma using decision tree analysis, J. Thorac. Oncol. : official publication of the International Association for the Study of Lung Cancer 11 (4) (2016) 573–582.
[35] J. Kang, Y.J. Choi, I.K. Kim, H.S. Lee, H. Kim, S.H. Baik, et al., LASSO-based machine learning algorithm for prediction of lymph node metastasis in T1 colorectal cancer, Cancer research and treatment 53 (3) (2021) 773–783.
[36] J. Alderden, G.A. Pepper, A. Wilson, J.D. Whitney, S. Richardson, R. Butcher, et al., Predicting pressure injury in critical care patients: a machine-learning model, Am. J. Crit. Care 27 (6) (2018) 461–468, an official publication, American Association of Critical-Care Nurses.
[37] L. Hu, M. Chen, H. Dai, H. Wang, W. Yang, A metabolism-related gene signature predicts the prognosis of breast cancer patients: combined analysis of high-throughput sequencing and gene, Chip Data Sets 24 (4) (2022) 803–822.
[38] F. Liu, Y. Huang, F. Liu, H. Wang, Identification of immune-related genes in diagnosing atherosclerosis with rheumatoid arthritis through bioinformatics analysis and machine learning, Front. Immunol. 14 (2023) 1126647.
[39] Y.X. Zhang, J. Lv, J.Y. Bai, X. Pu, E.L. Dai, Identification of key biomarkers of the glomerulus in focal segmental glomerulosclerosis and their relationship with immune cell infiltration based on WGCNA and the LASSO algorithm, Ren. Fail. 45 (1) (2023) 2202264.
[40] A. Liberzon, A description of the molecular signatures database (MSigDB) web site, Methods Mol. Biol. 1150 (2014) 153–160.
[41] Y. Gao, L. Guan, R. Jia, W. Xiao, Y. Han, Y. Li, et al., High expression of PPFIA1 in human esophageal squamous cell carcinoma correlates with tumor metastasis and poor prognosis, BMC Cancer 23 (1) (2023) 417.
[42] G. Sun, H. Ye, H. Liu, T. Li, J. Li, X. Zhang, et al., ZPR1 is an immunodiagnostic biomarker and promotes tumor progression in esophageal squamous cell carcinoma, Cancer Sci. 115 (1) (2024) 70–82.
[43] N.R. Choi, W.G. Choi, A. Zhu, J. Park, Y.T. Kim, J. Hong, et al., Exploring the therapeutic effects of atractylodes macrocephala koidz against human gastric cancer, Nutrients 16 (7) (2024).
[44] D. Wang, R. Li, J. Jiang, H. Qian, W. Xu, Exosomal circRNAs: novel biomarkers and therapeutic targets for gastrointestinal tumors, Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 157 (2023) 114053.
[45] Z. Zhang, L. Zhu, Y. Ma, B. Wang, C. Ci, J. Zhang, et al., Study on the characteristics of intestinal flora composition in gastric cancer patients and healthy people in the qinghai-tibet plateau, Appl. Biochem. Biotechnol. 194 (4) (2022) 1510–1526.
[46] S.B. Amatya, S. Salmi, V. Kainulainen, P. Karihtala, J. Reunanen, Bacterial extracellular vesicles in gastrointestinal tract cancer: an unexplored territory, Cancers 13 (21) (2021).
[47] Q. Qiu, Y. Lin, Y. Ma, X. Li, J. Liang, Z. Chen, et al., Exploring the emerging role of the gut microbiota and tumor microenvironment in cancer immunotherapy, Front. Immunol. 11 (2020) 612202.
[48] S. Wang, S. Wang, Z. Wang, A survey on multi-omics-based cancer diagnosis using machine learning with the potential application in gastrointestinal cancer, Front. Med. 9 (2022) 1109365.
[49] N. Gilani, R. Arabi Belaghi, Y. Aftabi, E. Faramarzi, T. Edgünlü, M.H. Somi, Identifying potential miRNA biomarkers for gastric cancer diagnosis using machine learning variable selection approach, Front. Genet. 12 (2021) 779455.
[50] Z.J. Zheng, Y.S. Li, J.D. Zhu, H.Y. Zou, W.K. Fang, Y.Y. Cui, et al., Construction of the six-lncRNA prognosis signature as a novel biomarker in esophageal squamous cell carcinoma, Front. Genet. 13 (2022) 839589.
[51] H. Zheng, X. Han, Q. Liu, L. Zhou, Y. Zhu, J. Wang, et al., Construction of immune-related molecular diagnostic and predictive models of hepatocellular carcinoma based on machine learning, Heliyon 10 (2) (2024) e24854.
[52] C. Meng, C. Bai, T.D. Brown, L.E. Hood, Q. Tian, Human gut microbiota and gastrointestinal cancer, Dev. Reprod. Biol. 16 (1) (2018) 33–49.
[53] C.C. Wong, J. Yu, Gut microbiota in colorectal cancer development and therapy, Nat. Rev. Clin. Oncol. 20 (7) (2023) 429–452.
[54] X. Liu, G. Zhang, S. Li, Y. Liu, K. Ma, L. Wang, Identification of gut microbes-related molecular subtypes and their biomarkers in colorectal cancer, Aging 16 (3) (2024) 2249–2272.
[55] A. Antonell, A. Lladó, R. Sánchez-Valle, C. Sanfeliu, T. Casserras, L. Rami, et al., Altered blood gene expression of tumor-related genes (PRKCB, BECN1, and CDKN2A) in alzheimer's disease, Mol. Neurobiol. 53 (9) (2016) 5902–5911.
[56] G. Hu, J. Du, B. Wang, P. Song, S. Liu, Comprehensive analysis of the clinical and prognostic significance of SFRP1 and PRKCB expression in non-small cell lung cancer: a retrospective analysis, Eur. J. Cancer Prev. : the official journal of the European Cancer Prevention Organisation (ECP) 33 (1) (2024) 45–52.
[57] J. Wang, M. Shi, H. Zhang, H. Zhou, Z. Huang, Y. Zhou, et al., PRKCB is relevant to prognosis of lung adenocarcinoma through methylation and immune infiltration, Thoracic cancer 13 (12) (2022) 1837–1849.
[58] J.H. Fan, M.M. Xu, L.M. Zhou, Z.W. Gui, L. Huang, X.G. Li, et al., Integrating network pharmacology deciphers the action mechanism of Zuojin capsule in suppressing colorectal cancer, Phytomedicine : international journal of phytotherapy and phytopharmacology 96 (2022) 153881.
[59] K. Daniunaite, A. Bakavicius, K. Zukauskaite, I. Rauluseviciute, J.R. Lazutka, A. Ulys, et al., Promoter methylation of PRKCB, ADAMTS12, and NAALAD2 is specific to prostate cancer and predicts biochemical disease recurrence, Int. J. Mol. Sci. 22 (11) (2021).
[60] M.A. Jarzabek, W.R. Proctor, J. Vogt, R. Desai, P. Dicker, G. Cain, et al., Interrogation of transcriptomic changes associated with drug-induced hepatic sinusoidal dilatation in colorectal cancer, PLoS One 13 (6) (2018) e0198099.
[61] D. Yang, Y. Zhou, Y. Zhang, Y. Su, J. Shen, B. Yu, et al., Comprehensive analysis of scRNA-Seq and bulk RNA-Seq data reveals dynamic changes in tumor-associated neutrophils in the tumor microenvironment of hepatocellular carcinoma and leads to the establishment of a neutrophil-related prognostic model, Cancer Immunol. Immunother. : CII. 72 (12) (2023) 4323–4335.
[62] R. Jia, S. Liu, J. Xu, X. Liang, IL16 deficiency enhances Th1 and cytotoxic T lymphocyte response against influenza A virus infection, Bioscience trends 13 (6) (2020) 516–522.
[63] L. Milke, K. Schulz, A. Weigert, W. Sha, T. Schmid, B. Brüne, Depletion of tristetraprolin in breast cancer cells increases interleukin-16 expression and promotes tumor infiltration with monocytes/macrophages, Carcinogenesis 34 (4) (2013) 850–857.
[64] W. Fang, C. Shi, Y. Wang, J. Song, L. Zhang, microRNA-128-3p inhibits CD4+ regulatory T cells enrichment by targeting interleukin 16 in gastric cancer, Bioengineered 13 (1) (2022) 1025–1038.
[65] C.K. Fu, M.C. Mong, H.E. Tzeng, M.D. Yang, J.C. Chen, T.C. Hsia, et al., The significant contribution of interleukin-16 genotypes, smoking, alcohol drinking, and Helicobacter pylori infection to gastric cancer, In vivo (Athens, Greece) 38 (1) (2024) 90–97.